Calculation of Errors in Rates

The formulas presented here provide a means of estimating the confidence interval around a single rate and for determining whether the difference between two rates is statistically significant.

***Definition:** A *confidence interval* is a range above and below an observed rate within which we would expect the "true" rate to lie a certain percentage of the time (usually 95%).

Calculation of a confidence interval recognizes that an observed rate is not a precise estimate of the underlying rate because the observed rate is influenced by random error. The formulas below are exactly the same as the ones used for a random sample from a larger population. The population rate for a given year based on a complete count can be considered a sample of one of a large number of possible measurements, all of which cluster in a normal distribution (bell curve) around the "true" (unknown) rate of the population. The larger the numerator of the measured rate, the better the rate will estimate the true or underlying rate of the population. The confidence interval accounts for only random measurement error. Systematic errors or biases in measurement may still be present and cannot be assessed by these formulas.

These formulas apply to any proportion or simple (crude) rate. Random errors may also be estimated for adjusted rates and other more complex measures, but a description of this is beyond the scope of the present Primer.

Proportions vs. Percentages vs. Rates

The formulas below are expressed in terms of p, or the proportion or fraction of a population that has a certain characteristic (e.g., death, low birthweight, early prenatal care). In this context, the terms proportion, percentage, and rate are used interchangeably.
For example, in 1995 Wake County had a resident population of approximately 518,000 out of which approximately 2,900 died during the year. The proportion who died is 2,900 / 518,000 or .005598. For the percentage who died, multiply by 100; the result is .5598%. A percentage is simply a rate per 100. For a rate per 1,000, multiply the proportion by 1,000; the result is 5.598 deaths per 1,000 population. The number of deaths per 100,000 is 559.8. So the multiplier is completely arbitrary, though for rare events we usually use 1,000 or higher so that the rate is not a decimal fraction.

The formulas presented below use p, or the proportion, so a percentage or rate has been converted back to the proportion (by dividing by the multiplier) in these examples.

Infant Death Rates

The infant death rates (expressed per 1,000 live births) reported in State Center for Health Statistics publications are not strictly proportions since the deaths and births occurred during a particular calendar year. Though approximately one-half of infant deaths occur on the first day of life, some of the infant deaths that occur in a given year are to babies born in the previous calendar year. Technically, the more correct way to compute the proportion of babies who before their first birthday would be to use a linked birth/infant death file to track a population of births (also called a birth cohort) through the first year of life. But in practicality this difference is small. We suggest that the formulas below may reasonably be used for infant deaths rates reported as usual based on year of occurrence and expressed as the proportion of babies who die.

✐ *Formula*:

$$\text{Infant death rate} = \frac{\text{\# deaths under 1 year of age}}{\text{\# of live births}} \times 1000$$

Confidence Intervals

We can compute a confidence interval around a proportion or rate. The confidence interval is the interval within which we would expect the "true" rate to fall a certain percentage of the time. A 95% confidence interval is frequently used, which means using a multiplier ("Z" value) of 1.96. For a 99% confidence interval, one would use the multiplier 2.57. Let us say that in hypothetical Rocky County there are 20 infant deaths (d) out of a population of 1,900 live births (n) in a single year. The proportion dying (p) is 20 / 1,900 = .0105. You can also say that 1.05 percent died or that the infant death rate is 10.5 per 1,000 births for that year.

✐ *Formula:*

$$95\%\ \text{Confidence Interval} = p \pm 1.96\sqrt{\frac{p\,q}{n}}$$